SOLUTION BRIEF

Optimize and Secure AI Workloads and LLMs with FortiAIGate

Executive Summary

Organizations worldwide are rapidly adopting private AI and large language model (LLM) technology to accelerate innovation and stay competitive. Nearly every employee is using AI for something, yet there are no standards or regulations. This means that security teams are playing whack-a-mole with the exposure it creates, finding it difficult to protect AI workloads while ensuring security controls don't slow down business.

FortiAIGate addresses the new security, governance, and operational challenges that arise when organizations deploy AI models and agents. The solution ensures that sensitive data isn't leaked and that model outputs remain compliant and safe. Further, it provides a standardized way to connect applications, APIs, and AI models, regardless of the provider or deployment location.

69%

of CISOs say their teams lack the necessary skills to work with GenAI technology.[1]



Enforce Security Guardrails and Get Visibility into AI Workloads

For an app to interact with LLMs securely and efficiently, organizations need intelligent routing, consumption control, and security enforcement. In addition, AI changes the game by embedding ML training and inference within the app layer, increasing business-logic complexity. When data flows, blending classic and AI-enhanced layers, more blind spots emerge.

FortiAIGate addresses the key AI challenges of controlling and securing model interactions and the lack of visibility and governance. This includes protecting against threats such as prompt injection, data exfiltration, and model misuse by applying runtime guardrails and policies to every AI request and response.

As enterprises adopt multiple AI models across clouds, departments, and vendors, it becomes difficult to track who is using what, with which data, and under what policies. The AI gateway centralizes monitoring, auditing, and access control, giving organizations unified visibility and enforceable governance across distributed environments.

In addition, FortiAIGate provides a standardized way to connect applications, APIs, and AI models, regardless of the provider or deployment location. This consistency simplifies management, enables compliance enforcement at scale, and allows security and DevOps teams to extend existing cloud or API protection frameworks to the AI layer.

Optimize and Protect AI Investments

In a nutshell, FortiAIGate covers the AI runtime attack surface, protecting LLMs from security threats. Deployed between the app and the model, it manages LLM traffic flows and applies guardrails on the model input and output. With FortiAIGate, you can optimize and secure your AI investments and intellectual property by detecting and preventing model manipulation, prompt injection, excessive consumption, and data leakage.

The local, containerized deployment of FortiAIGate offers a lightweight footprint, scalability, and intelligent LLM traffic steering, acting as a proxy between the app and the model to optimize delivery. Output caching offloads duplicate tasks from the model, reducing token consumption and compute costs, while improving the user experience. The solution applies customized guardrails to each LLM endpoint to accurately detect prompt injection, jailbreaking, model poisoning, and attempts to consume excessive resources. Our context-aware DLP engine monitors inbound and outbound traffic to prevent a data breach or model extraction.

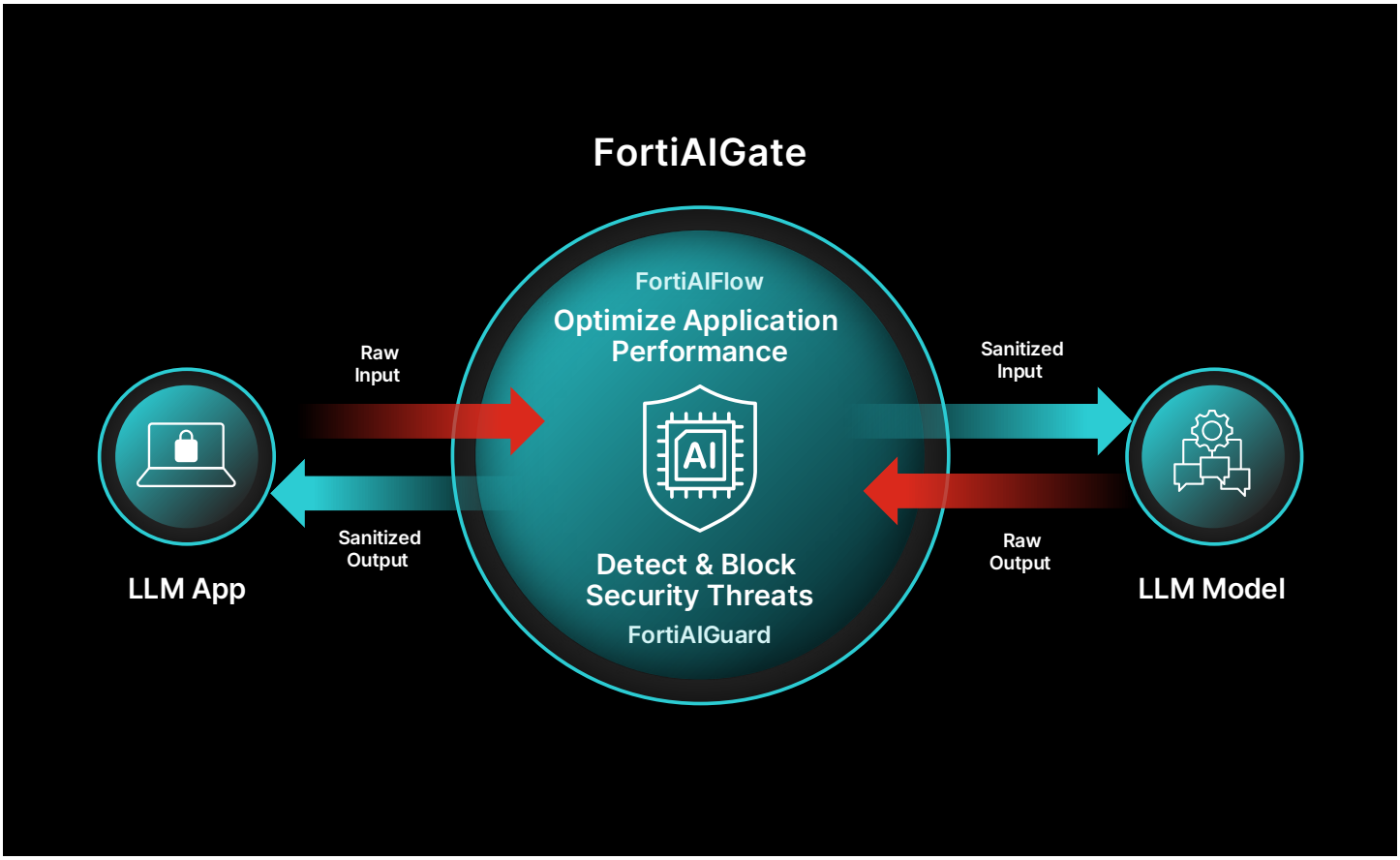


Figure 1: FortiAIGate delivers AI runtime security to defend LLMs.

FortiAIGate helps security teams with:

- **Easy management:** Built-in GUI, no manual YAML configuration or Python coding
- **Complete visibility:** Full coverage and clear mapping of OWASP Top 10 risks for LLM applications
- **Extensive LLM proxy and security features:** Includes API gateway and DDoS protection
- **Scalable and flexible deployment:** Deployed as a built-in Kubernetes cluster, auto-scales with LLM utilization
- **High performance:** Multi-GPU and SmartNIC-accelerated for low-latency proxy offloading

Confidently Innovate

By securing your investments in applications using LLMs, you can innovate with confidence. FortiAIGate protects against emerging AI threats while driving efficiency, governance, and cost optimization.

Reduce threat exposure

By enforcing guardrails and policies at every model interaction, FortiAIGate helps achieve governance and reduce AI risk exposure. It prevents the exfiltration of sensitive data, mitigates prompt-based attacks, and ensures responsible AI outputs that align with compliance and ethical standards. This governance extends across distributed AI environments—cloud, on-premises, and edge—providing unified visibility, auditing, and control.

Accelerate AI adoption

At the same time, FortiAIGate increases operational efficiency by seamlessly integrating into existing infrastructure. Intelligent routing capabilities automatically direct AI queries to the optimal model based on cost, latency, and performance metrics. This enables teams to safely scale AI experimentation and deployment without introducing unmanaged risk or complexity.

Control costs

Finally, the platform optimizes AI cost management and performance through advanced traffic steering, caching, and workload optimization. It helps organizations reduce unnecessary API calls, control model usage costs, and maintain predictable performance across workloads.

With FortiAIGate, businesses can achieve:

- Proactive defense against LLM-specific attacks and misuse
- Centralized governance and audit and compliance enforcement
- Intelligent model routing and performance tuning
- Reduced operational overhead and AI spend

Enable Secure AI Adoption

FortiAIGate protects AI workloads from evolving threats while maintaining business agility. With FortiAIGate, you can confidently scale AI initiatives, reduce operational complexity, and secure your data and intellectual property. The result is a future-ready foundation for innovation, where security and performance go hand in hand to unlock the full potential of enterprise AI.

[1] NTT DATA Research Reveals C-Suite Misalignment Over GenAI Adoption, NTT Data, June 20, 2025. https://www.nttdata.com/global/en/news/press-release/2025/june/062000

Copyright © 2025 Fortinet, Inc. All rights reserved. Fortinet®, FortiGate®, FortiCare® and FortiGuard®, and certain other marks are registered trademarks of Fortinet, Inc., and other Fortinet names herein may also be registered and/or common law trademarks of Fortinet. All other product or company names may be trademarks of their respective owners. Performance and other metrics contained herein were attained in internal lab tests under ideal conditions, and actual performance and other results may vary. Network variables, different network environments and other conditions may affect performance results. Nothing herein represents any binding commitment by Fortinet, and Fortinet disclaims all warranties, whether express or implied, except to the extent Fortinet enters a binding written contract, signed by Fortinet's SVP Legal and above, with a purchaser that expressly warrants that the identified product will perform according to certain expressly-identified performance metrics and, in such event, only the specific performance metrics expressly identified in such binding written contract shall be binding on Fortinet. For absolute clarity, any such warranty will be limited to performance in the same ideal conditions as in Fortinet's internal lab tests. Fortinet disclaims in full any covenants, representations, and guarantees pursuant hereto, whether express or implied. Fortinet reserves the right to change, modify, transfer, or otherwise revise this publication without notice, and the most current version of the publication shall be applicable.

November 11, 2025 2:02 PM / MKTG-2492-0-0-EN